Reuse Library Interoperability and the World Wide Web

Shirley V. Browne
University of Tennessee Netlib Development Group
browne@cs.utk.edu

James W. Moore
Chairman of the Reuse Library Interoperability Group (RIG)
moorej@acm.org

October 4, 1996

1 Prologue

Shortly after the collapse of Soviet domination in Eastern Europe, a Polish housewife was interviewed by National Public Radio. She complained that the introduction of free enterprise to Poland was a great inconvenience. Now, instead of shopping at the most conveniently located store, secure in the knowledge that prices were uniform, she had to go from store to store comparing prices, in order to be sure that she was making the best buy. "Why should prices be different at each store?" she complained. "After all, they are all selling the same things." On the surface, her remarks seem to make a great deal of sense. If we can find the fallacy in her thinking, we will learn a small lesson about a very important subject, free-market economics, and more relevantly, a big lesson about a less important subject, software reuse libraries.

2 Introduction

The Reuse Library Interoperability Group (RIG) was formed in 1991 for the purpose of drafting standards enabling the interoperation of software reuse libraries. At that time, prevailing wisdom among many reuse library operators was that each should be a stand-along operation. Many operators saw a need for only a single library, their own, and most strived to provide the most general possible services to appeal to a broad community of users. The ASSET program, initiated byt he Advanced Research Projects Agency STARS program, was the first to make the claim that it should properly by one part of a network of interoperating libraries [8]. Shortly thereafter, the RIG was formed, initially as a collaboration between the STARS program and the Air Force RAASP program, but growing within six months to a self-sustaining cooperation among twelve chartering organizations. The RIG has grown to include over twenty members from government, industry, and academic reuse libraries. It has produced a number of technical reports and proposed interoperability standards, some of which are described in this report.

3 Fundamental Assumptions about Reuse Libraries

When the RIG was formed, some of its fundamental claims regarding the nature of a reuse industry were regarded as questionable, controversial or nonsensical, even by some reuse library operators. Since then the explosive growth in the usage of the Internet triggered by the World Wide Web (WWW) has caused a sea of change in attitudes regarding usage of the Internet. Far from controversial, many of the RIG's five-year-old claims are now taken for granted, even considered "old hat". Let's take a look at some of them [6, 9].

Interoperable libraries

The status quo in 1991 was that each reuse library was a stand-along entity attempting to provide all needed services to all potential users. The (unstated but apparent) basis for competition was that each library felt that it was – or could be – or should be – the single best reuse library, that everyone should use it, and that there was no need for others. Since there would be only a single library, the subject of interoperability was moot. The World Wide Web itself provides a conclusive counterexample to this kind of thinking. Users of the Web now routinely benefit from interoperation when a home page smoothly incorporates links to other pages on other machines, even in other countries.

Data models to permit the sharing of information

The first standard proposed by the RIG was a data model describing exchangeable information regarding reusable assets. The criticism was "Why not a concrete format; why the additional level of abstraction?" Well, an abstract model provides for wide use and implementation of the standard in various environments and on various platforms with widely varying capabilities. Today, any user of a Web browser benefits from the fact that HTML (HyperText Markup Language) prescribes a model for documents on the Web, rather than a presentation format, allowing both graphical and non-graphical browsers to obtain the information embedded in the documents.

User interface independent of the catalog

In 1991, most of the major reuse libraries bundled a user interface into the monolithic set of services provided. If you wanted to use their library, you had to learn their interface. Today, Web users take it for granted that they can use Netscape, Mosaic, Lynx, or a variety of other browsers to navigate the Web. Of course, it's the existence of a document data model that makes this possible.

Catalogs **without** *components*

Many observers found it perplexing in 1991 that the RIG was concerning itself with the concept that cataloguing components was an activity distinct from storing them and that catalogues could effectively refer to components stored in other libraries. Users of the Web have now become quite familiar with the concept that a particular home page can refer to documents or images stored on machines that are quite far away and under a different administrative domain.

Value-added services distinct from libraries

Operators of monolithic libraries thought in terms of a monopoly on providing services to their users. After all, the only route to their products and services was via their user interface and their catalog. Users of the Web now know that the easy ability to reference documents held by others permits interested individuals and even entrepreneurs to add value to existing libraries by creating catalogs to collections held by others without actually storing the referenced components on the same machine as the catalog. Furthermore, search tools and services are widely available, some for a fee.

Marketplace

The most radical concept in 1991 was that a network of interoperating reuse libraries could provide a marketplace where users, producers and providers of various value-added services could interact to achieve a broad variety of commercial goals. The explosive growth of commercial enterprise on the Web is powerful evidence of the demand for commercial marketplaces. Entrepreneurs are solving the problems of performing commercial transactions on the Web. Today, transactions to license, pay for, and download commercial software are routinely carried out over the Web. Web-based reuse services and products are just beginning to appear, but we expect this market to grow considerably within the next few years.

4 Benefits of Interoperability Standards

The RIG has identified three beneficiaries of interoperability standards [6]:

- Users: Users of reuse libraries can access products and services from multiple reuse libraries via a single interface.
- Libraries: Libraries can reduce the burden of storing redundant contents by simply pointing to components already stored in other libraries.
- Community: The value-added functions of any library would be available to any user, creating niches for entrepreneurs and enabling a software reuse industry.

By focusing on the nature of these benefits, the RIG obtains guidance for determining appropriate candidate areas for standardization.

Benefit to Users

The key to permitting users of one library to access the contents of other libraries is to abstract the data to be interchanged apart from any concrete format for representation or any specific method for presentation. Of course, HTML provides the analogous function for the World Wide Web. The HTTP protocol used by the WWW is not the only possibility for interconnecting reuse libraries. Alternatives include the X.500 Directory Services protocol and the Z39.50 Information Retrieval protocol. Ideally, it should be possible to access reuse library resources over different types of underlying networks and network protocols. For this reason, the RIG chooses to abstract its data model away from the selection of a specific protocol. This decision leads to a three-level framework:

1. Data Model: The data model abstractly describes the information to be interchanged among reuse libraries. An example is the RIG Basic Interoperability Data Model (BIDM) [1].

2. Format: A particular format explains how an instance of a data model is concretely represented. Examples are the RIG HTML and SGML binding of the BIDM described in section 6.

3. Protocol: A protocol describes the "conversation" between two nodes in a network in order to interchange information. Examples are HTTP and Z39.50.

Much of the work of the RIG has been involved with defining and elaborating the Basic Interoperability Data Model and with studying ways in which the data model might be "bound" to various formats and protocols. Now that specific bindings are being implemented, users clearly benefit from use of a common interface, such as a Web browser, to view the catalogs of various reuse libraries.

Benefit to Libraries

Our second claimed benefit of interoperability is to reduce the redundant holdings among reuse libraries. If libraries are able to interoperate by exchanging catalog records, and even mirroring each others software files, then a user of any one library can obtain goods and services offered by other libraries, and libraries will not have to duplicate holdings in order to offer comprehensive services to their users. Instead, each library can focus on tailoring its contents and services to a particular application domain or community of users.

A key to eliminating redundancy and maintaining version control between interoperating libraries is a standard for unique identifiers of reusable assets. The idea of a unique identifier is simple. Whenever a potentially reusable component is first published, a unique identifier is assigned to it. Any other component would have a different identifier. Any change to a component would require assignment of a distinct identifier. By observing these policies, one can ensure the vital property that if two components have the same identifier, then they have identical contents. This property can be used by a library to determine if a candidate for its collection is identical to one of its current holdings or identical to one held by an interoperating library.

Although the RIG has not yet adopted a standard for a unique identifier, it has reserved a place for one in the BIDM. The use of public key cryptography for ensuring the desired properties of the Unique ID is proposed in [10]. The Netlib Development Group, an organizational member of the RIG, has made a detailed proposal for how unique identifiers might be assigned to all resources available from the National HPCC Software Exchange [4].

The BIDM and unique identifiers are not the only concerns of the RIG. The BIDM is intended to define the *minimum* information that reuse libraries should be prepared to interchange. Different libraries have different software evaluation and certification policies which are specific to their unique missions. The RIG Asset Certification Framework (ACF) extension to the BIDM provides a common basis for exchanging and understanding different policies as well as certification activities and results. Another area in which library differ is in the management of intellectual property rights. Similar to the ACF, the RIG Intellectual Property Rights Framework (IPRF) allows libraries to describe and exchange their different rights management policies and procedures, as well as the results of carrying out these procedures. A library may have additional metadata, beyond that specified in the BIDM or its extensions, that it would like to make available, and it may wish to extend the BIDM for this purpose. Because it is expected that groups outside the RIG will define extensions to the basic model , the RIG is working on a formal meta-model that provides a consistent methodology for defining model extensions.

The RIG does not simply create standards for the use of its own members, although many of the major reuse libraries are members. Selected RIG products progress to the status of national standards via a Memorandum of Understanding with the IEEE Computer Society. Under the terms of the MOU, the RIG is given the "turf" for drafting standards in the area of reuse library interoperation. In return, the RIG agrees to participate in the IEEE balloting process which is used to promote a proposed RIG standard to the status of an IEEE standard.

Benefit to the Community

The final benefit is a more global one, the benefit to the entire community of suppliers and consumers. The availability of convenient interchange of data among libraries permits entrepreneurs to provide value-added services that may appeal to various users. The availability of a commercial marketplace is the economic incentive encouraging such developments. At least one such service is provided by any library – the value-added service of *selection*. Different libraries have different policies for adding items to their holdings. Knowledge of those criteria can be useful to consumers.

Consider, for example whether you would prefer to purchase a business suit at Brooks Brothers or at K-Mart. You know that they have differing goals in selecting their merchandise, and knowledge of those goals helps you decide where to shop. Accordingly, the selection policies applied by different reuse libraries may serve to establish an image that will help reusers as they shop for components.

As previously mentioned, the ability to add value via selection is not confined to the library operator. Independent entrepreneurs can build indices of components actually stored in many other libraries. The organization that they superimpose upon the selected components may have important value.

Another possibility for added value is the bundling of advice, service or support with a component. One step further would be the coupling of consulting services with families of components that serve to define an architecture.

Some catalogs might distinguish themselves by applying various certification criteria to ensure some measure of quality for components within different collections. Such certification criteria could be described and presented to users by means of the RIG Asset Certification Framework.

It's important to note that the commercial aspect of a marketplace can provide acceptable solutions to seemingly insolvable technical problems. The application of economic incentives is often effective in providing a non-technical solution to a problem. For example, in the area of liability, one could attempt to solve difficult technical problems to provide high confidence assurance of the quality of the components in one's catalog. On the other hand, one could adopt an economic approach to the problem by coupling readily available medium confidence assurance mechanisms with some sort of indemnification, such as a money-back guarantee. Such rights policies can be described in terms of the RIG Intellectual Property Rights Framework.

The importance of this putative benefit to communities is to remind us of the value of restraint. Careless standardization could have the unintended effect of making it unnecessarily difficult, or even unfeasible, to provide a particular value-added service. For example, a presumption embedded in a data model that component classification is faceted might have the effect of making it difficult to provide a navigational classification scheme in a domain-oriented library. As another example, prescribing the use of a particular protocol for catalog record interchange would prevent taking advantage of new Web protocols and technologies. The general lesson here is that the RIG should propose the *minimum* amount of standardization necessary to accomplish its objectives and to achieve the cited benefits. Furthermore, the RIG should capitalize upon and further cultivate its broad base of membership (government, academia, users, vendors) to ensure that many viewpoints are represented.

In all of these examples of value-added services we see that value is added by *differentiating* one's service from available alternatives, not by copying them. One classification suggests that reuse libraries will move into roles characterized as *local* or *organizational* libraries, dedicated to serving an affiliated group of users, *domain-specific* libraries, dedicated to serving a *keiretsu* of organizations working on related problems, and *reference* libraries, dedicated to helping users find libraries of interest and to tracing pedigrees and intellectual property relationships among components.

The incentive to differentiate is one of the great benefits of an open, robust, free market. An entrepreneur can "test the market" in offering different services, and the economic mechanisms of the marketplace will provide prompt indication of the value perceived by potential customers. When the RIG was formed, nobody envisioned the extent to which the Internet would become a part of the routine life of people who aren't computer-oriented professionals. Its incredible growth and easy connectivity now provide an immense marketplace as well as a number of important example for the RIG and other organizations.

The widespread adoption of interoperability standards, such as those proposed by the RIG, will enable reuse libraries to enter this marketplace and, in turn, enable them and the value-adding entrepreneurs to differentiate themselves. This will provide the basis for a robust, economically viable industry in software reuse. Even a small niche in a huge market is much more economically attractive than a big niche in a tiny market.

The remainder of this paper describes the various standards that have been proposed by the RIG, as well as an example of the use of these standards by the National HPCC Software Exchange, a network of interoperating HPCC software repositories.

5 The Basic Interoperability Data Model

The Basic Interoperability Data Model (BIDM), which is an IEEE standard (1420.1) for software reuse libraries, specifies a minimal set of metadata that a reuse library should provide about its reusable assets in order to interoperate with other reuse libraries [1]. The BIDM is expressed in terms of an extended entity-relationship data model that defines classes for assets (the reusable entities), the individual elements making up assets (i.e., files), libraries that provide assets, and organizations that develop and manage libraries and assets. The model was derived from careful study and negotiation of the commonalities between existing academic, government, and commercial reuse libraries, by representatives from these libraries. Reuse libraries need not adopt the BIDM internally, although many have. They can continue to use internal search and classification mechanisms appropriate to their unique missions while using the BIDM as a uniform external interface.

The BIDM may be visualized using the graphic notation of James Rumbaugh's *Object-Oriented Modeling and Design* [12]. Figure 1 provides a legend for the graphic notation. A pictorial view of the BIDM is shown in Figure 2.

A subclass inherits all attributes and relationships of its parent class. For example, the Asset, Element, Library, and Organization classes all inherit the Name attribute from the RIGObject class. The basic model may be extended by defining additional subclasses, as described in section 7.

Each of the classes, attributes, and relationships has a well-defined semantics which is specified in the BIDM document. The datatype and the allowed multiplicity for each attribute are also specified.

The semantics of the UniqueID attribute for the Asset and Element classes say that it is a globally unique name used to refer to an object – e.g., for the purpose of retrieving metadata about the object or of retrieving the object itself. Global uniqueness requires that no two objects be given the same UniqueID, even object owned by different repositories. Although various proposals for the

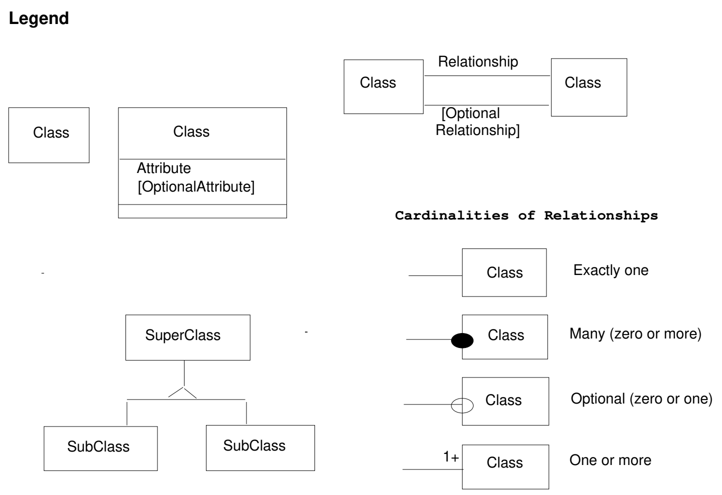

Figure 1: Legend for Data Model Notation

UniqueID attribute have been brought before the RIG, as yet none has been adopted. For lack of a better solution, most BIDM implementors are currently usin URLs for the UniqueID field. The RIG is monitoring progress by the Internet Engineering Task Force (IETF) on Uniform Resource Names (URNs) and may adopt URNs for the UniqueID attribute if and when URNs become a standard.

Many organizations and disciplines use controlled vocabularies for one or more of the BIDM attributes, such as Domain and Keyword. For example, several mathematical software repositories and companies use the Guide to Mathematical Software (GAMS) to classify mathematical software [2]. As described in section 7, work is underway on a model that would allow a library to indicate that is using a particular controlled vocabulary for a particular attribute.

6 Data Model Bindings

In order for catalog information to be exchanged between software repositories, the abstract data model described in the preceding section is mapped to

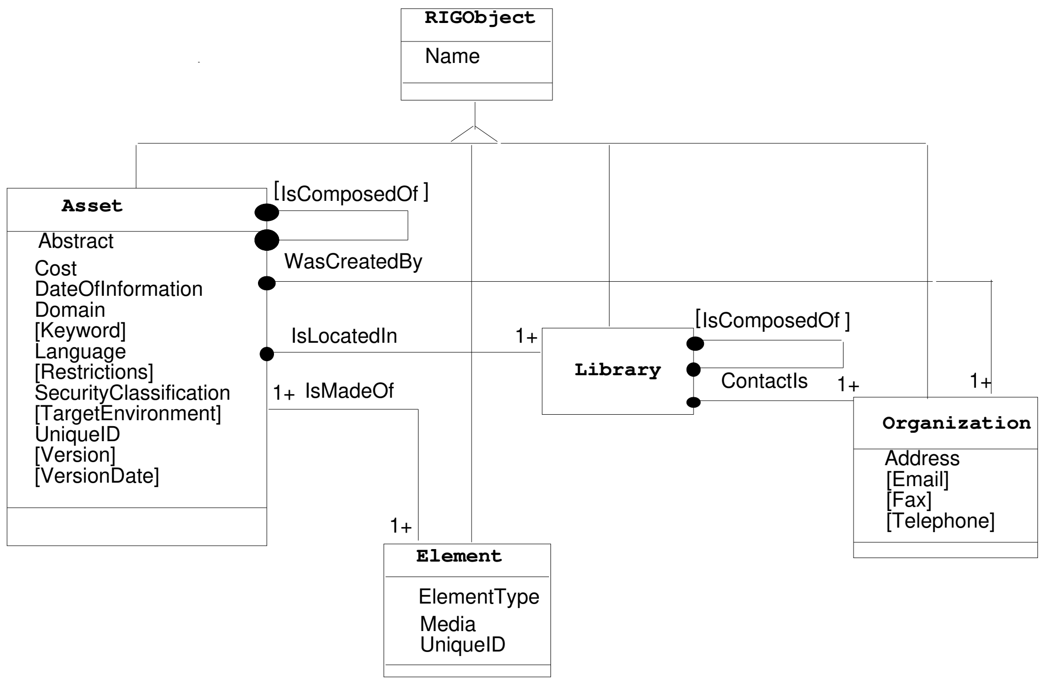

Figure 2: Basic Interoperability Data Model

a concrete syntax that can be transferred using a file transfer protocol such as FTP or HTTP. So far the RIG has defined two such bindings, one that maps the BIDM to an SGML Document Type Definition (DTD), and another that maps the BIDM to META and LINK tags in the header of an HTML document. These bindings have been implemented and tested by RIG members. A document containing the specifications for these bindings is currently in the IEEE standardization process.

The SGML DTD for the RIG BIDM and the Asset Certification Framework extension and examples of its use are available from the RIG home page at `http://www.rig.org/`. Each class, attribute, and relationship is defined as an SGML element. Subclasses are represented by nesting the subclass element within the parent class element. For any of the relationships, the implementor has the option of nesting the destination class SGML elements within the parent class SGML element, or of specifying an identifier for a separate file which contains the metadata for the destination class.

With the HTML binding, the attribute and relationship metadata for an object is placed in META and LINK tags in the header of an HTML document that otherwise describes the object – for example an HTML page that describes a software asset or an organization. A detailed description of the HTML binding, as well as examples of its use, are available from the RIG home page at `http://www.rig.org/`. The value of a BIDM attribute may be given either by placing it in a META tag of the form

```
<META NAME="data-model.class.attribute" CONTENT="value">
```

or by placing the value in a separate file and pointing to it using a LINK tag of the form

```
<LINK REL="data-model.class.attribute"
   HREF="URL of file containing the attribute value ">
```

The latter method is useful when the value contains special characters, such as quotations marks or HTML markup, that can't occur in the CONTENT of a META tag.

The metadata for an object that is the destination of a relationship may be in-lined in the HTML file that describes the source object, with metadata for multiple destinations correlated by means of a correlation number, by using META tags of the form

```
<META NAME="data-model.src-class.rel.dest-class[.cn].attribute">
```

where `rel` stands for relationship and `cn` for correlation number. Alternatively, if another file contains the metadata for the destination of a relationship, the file may be specified by using a LINK tag of the form

```
<LINK REL="BIDM.source-class.relationship.dest-class"
      HREF="URL for destination">
```

The binding process involves some form of collection to retrieve, parse, and validate metadata located in HTML or SGML files stored on the Internet or on an organization's internal Intranet. A typical scenario would be for a library administrator to initiate a Web spider to collect and validate metadata files from outside libraries. This metadata could then be incorporated into the library's environment where it could be stored in a database or directory structure that could be searched by the library's users.

Because the HTML and SGML bindings have been in use for less than a year, it is too earlier to tell which will end up being most widely adopted, or if a completely different binding, such as perhaps Z39.50, will prove more successful.

7 Model Extensions

Although the Basic Interoperability Data Model has greatly enhanced the ability of reuse libraries to interoperate, it is desirable to be able to extend the basic model to cover specific areas more thoroughly or to meet the needs of specialized libraries. One area for which an extension has already been defined is that of asset evaluation and certification. The extension is the RIG Asset Certification Framework, which defines a standard for the consistent structure, labeling, and description of evaluation and certification policies and results, and which is discussed further below. The RIG has defined another extension, also discussed below, called the Intellectual Property Rights Framework, which provides a consistent framework for labeling and describing intellectual property rights and other legal restrictions on software assets. Another reason extensions are needed is that a library may have additional metadata, beyond that specified in the BIDM, that it would like to make available, and it may wish to extend the BIDM for this purpose.

Because it is expected that extensions to the basic model will be defined by groups outside the RIG, and to ensure that the RIG itself follows a consistent methodology in defining model extensions, the RIG is working on a formal meta-model for describing allowed extensions. Although the short term goal for this meta-model is that it be understandable and usable by human data modelers, a longer term goal is that it be understandable by intelligent agent programs that would interpret and process metadata from the basic data model and its extensions automatically.

7.1 The Asset Certification Framework

Most software reuse libraries organize their evaluation and certification policies by levels. These levels provide a quick reference for the user in determining what evaluation and certification criteria have been met by particular assets. In general, increasing levels represent increasing confidence in the asset, as well as increasing certification effort and cost. However, each library has defined its levels differently, and the different levels and policies are confusing to users of multiple interoperating libraries. Each reuse library needs to be able to define certification policies that are unique to its particular mission and that are compliant with domain-specific standards. Rather than attempting to drive all libraries to a standard set of levels, the Asset Certification Framework (ACF) prescribes a standard for organizing and describing different policies. Thus, the ACF provides a common basis for comparing different policies and for understanding different libraries' evaluation and certification activities and results.

The ACF extends the BIDM by adding the AwardedWith relationship to the Asset class of the BIDM and by defining additional classes of objects that are relevant to evaluation and certification. A pictorial view of the ACF, with attributes of the original BIDM classes omitted, is shown in Figure 3. A tabular view of the ACF is shown in Figure 4.

Certification quality factors are high level evaluation criteria, such as completeness, correctness, and reliability. Certification properties define features or characteristics of an asset that may be assessed as being true or false, or that may be measured. Certification methods are documented evaluation techniques, which may include compilation, static analysis, inspection, testing, formal verification, and benchmarking.

7.2 Intellectual Property Rights Framework

The RIG Technical Committee on Intellectual Property Rights has developed a BIDM extension, called the Intellectual Property Rights Framework (IPRF), for labeling assets with information regarding legal restrictions commonly asserted in the United States, such as copyright, patents, licensing, and export restrictions. Similar to the Asset Certification Framework, the IPRF provides a common framework for interoperating libraries to describe and exchange their rights management policies. Similar to how asset certificates are linked to the certification policy that defined them and to the organization that did the certification in the ACF, rights assessments and licensing terms are linked to their defining policies and responsible organizations in the IPRF.

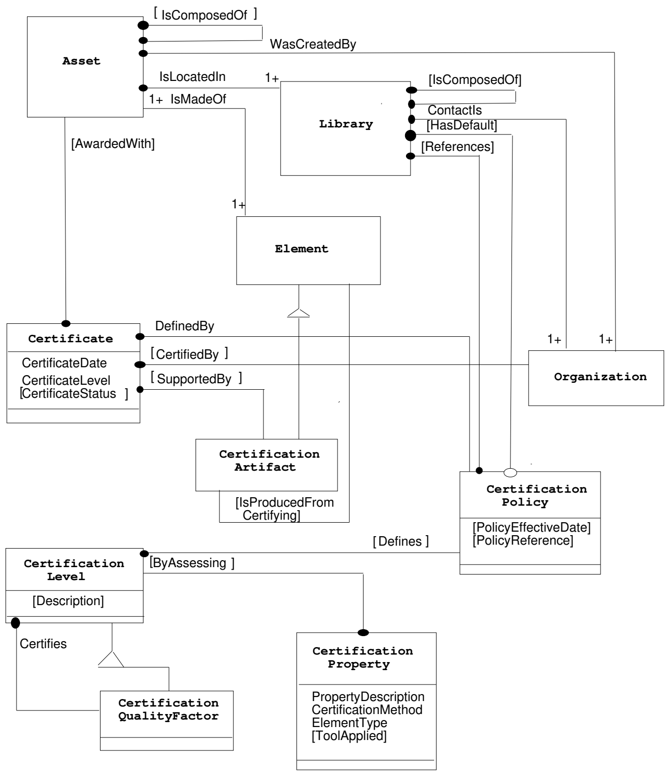

Figure 3: Asset Certification Framework

Certification Levels	Certification Property		Element Type
	Property Description	Certification Method	

Figure 4: Tabular View of Asset Certification Framework

The proposed RIG IPRF has been completed and is ready to submit to the IEEE balloting process. One goal of this work is to enable pre-negotiation of agreements between reuse libraries that facility large-scale sharing of restricted software – i.e., so that a separate agreement for each software asset and each pair of interoperating libraries does not need to be negotiated. The RIG hopes that by providing the means to unambiguously describe export and other legal restrictions on software, risks and fears of liability and litigation will be reduced and not unduly impede the exchange of software between libraries.

7.3 Meta-Model

The approach being taken by the RIG in defining a formal model for describing model extensions is to define the allowed extensions in terms of formal data modeling notation [12]. Data modelers will thus be able to determine unambiguously how new classes, attributes, and relationships may be defined, as well as how to represent these entities in terms of the same data modeling notation.

The BIDM makes no provision for controlled vocabularies. However, it is clearly desirable for reuse libraries to be able to use existing controlled vocabularies, such as keyword lists, taxonomies, and thesauri, as well as place other constraints on values of an attribute, such as a particular date format. To meet this need, the meta-model will include a scheme for describing constraints on the possible values of an attribute.

8 The National HPCC Software Exchange

The National HPCC Software Exchange (NHSE) provides a uniform interface to a distributed set of discipline-oriented HPCC repositories [5] As such, the NHSE is a *virtual repository*, in that it catalogs and points to software maintained elsewhere, except for archive and mirror copies stored on NHSE machines. A virtual repository is a type of interoperation that involves a hierarchical relationship.

In many cases, a discipline-oriented repository will wish to provide its own specialized interface to its software collection. The repository may use classification schemes and search tools tuned to its particular discipline. For example, the Netlib [7] and GAMS [3] mathematical software repositories use the GAMS classification scheme and are developing expert search subsystems for specific GAMS classes. Discipline-oriented repositories will also be in the best position to review and evaluate software within their own domains. In addition to providing access to its own software, a repository may wish to import software descriptions from other repositories and make this software available from its own interface. For example, a computational chemistry repository may wish to provide access to mathematical software and to parallel processing tools in a manner tuned to the computational chemistry discipline.

The NHSE is using the RIG BIDM bindings as its interoperability mechanism. Participating HPCC repositories and some individual contributors have placed META and LINK tags in the headers of HTML files that describe their software resources. Some repositories are making use of the SGML binding as well. In addition to the BIDM fields, the NHSE data model includes a few additional fields that are desirable for NHSE interoperation. The relevant data model for a field is currently specified by prefixing the field name with the data model name in the name attribute of the META tag. In the future, NHSE extensions to the BIDM will be described using the RIG meta-model which is currently under development. The NHSE is developing a toolkit called Repository in a Box (RIB) that will assist repository maintainers in creating and maintaining software catalog records, in exchanging these records with other repositories (including the top-level virtual NHSE repository), and in providing a user interface to their software catalog.

As a virtual repository, the NHSE sees a need for a globally unique identifier that unambiguously identifiers a particular version of a software asset. Such unambiguous identification is necessary for a number of reasons, including the following:

- version tracking
- associating testing and review metadata with the exact version that was reviewed
- reporting and reproducing scientific results

However, the NHSE also sees a need for a stable name for a resource that does not change every time there is a minor bug fix or revision. The NHSE is currently experimenting with using both URLs and URNs in the metadata that is exchanged using the Web bindings of the RIG BIDM. The NHSE data model includes an additional fingerprint field for identifying the exact version of a file. The fingerprint scheme currently used by the NHSE is MD5 [11].

Distributed maintenance of resources, although desirable for maintaining information close to its source and thus allowing local control and keeping it up-to-date, raises performance and reliability problems for access by remote users. Performance and reliability problems can be solved by replication and caching. However, replication and caching raise consistency and intellectual property rights issues.

The Resource Cataloging and Distribution System (RCDS) under development at the University of Tennessee uses a consistency model based on Location Independent File Names (LIFNs). Once assigned, a LIFN is immutably bound to a particular sequence of bytes. After updating a file, a publisher assigns it a new LIFN, registers the new URN-to-LIFN binding with an RCDS catalog server, and notifies authorized file servers who can then acquire the new file and notify a location server of the new LIFN-to-URL binding. Thus, the RCDS scheme is a combination of TTL-based "pull" consistency, with file servers pulling updates at their convenience, and invalidation-based "push" updating by efficient propagation of meta-information updates among catalog servers.

The NHSE is planning to mirror authorized copies of software from the various HPCC repositories and individual software providers on NHSE file servers. The NHSE is also planning to run experimental RCDS catalog and location servers on the distributed set of NHSE servers. Experiments will be carried out to compare the performance and efficiency of the RCDS file replication approach with other proposed replication and caching schemes.

The NHSE has designed a software review policy that enables easy access by users to information about software quality, but which is flexible enough to be used across and specialized to different disciplines. The three review levels recognized by the NHSE are the following: *Unreviewed*, *Checked*, and *Reviewed*. The *Unreviewed* designation means only that the software has been accepted into the owning repository and is thus within the scope of HPCC and of the discipline of that repository. The *Checked* designation means that the software has been checked by a librarian for conformance with the NHSE software guidelines. The *Reviewed* designation means that the software has been reviewed by an expert in the appropriate field. Domain-specific repositories and expert reviewers are expected to refine the NHSE software review policy by adding additional review criteria, evaluation properties, and evaluation methods and tools. The NHSE also provides for soliciting and publishing author claims and user comments about software quality. All software exported to the NHSE by its owning repository or by an individual contributor is to be tagged with its current review level and with a pointer to a review abstract which describes the software's current review status and includes pointers to supporting material. The review information is also encoded in terms of the RIG Asset Certification Framework for exchange with other software repositories.

Protection of intellectual property rights should not unduly impede or slow access to software. The NHSE is faced with the task of distributing and providing efficient access to HPCC software, some of which has access restrictions. The NHSE is currently undertaking a study of how efficient access can be provided while meeting legal restrictions and security objectives, and without exposing third parties, such as NHSE online service providers, to legal liability for rights infringement or violation of U.S. export law.

9 Epilogue

Now we know enough about economics to resolve the dilemma posed by the Polish housewife. Soviet Poland operated under a planned economy where a uniform set of goods were sold in all outlets at uniform prices. Post-Soviet Poland relaxed the requirement of identical prices, but the production and distribution system remaining from the previous regime still resulted in stores providing identical selections of identical goods. In short, there was no differentiation among the value of goods and services provided at the retail level. In this context, the complaint of the Polish housewife was justified. The value of a free market economy becomes apparent only when entrepreneurs justify their prices by adding value even if only through selection. This process results in the differentiation of their goods and services.

In 1991, many of the major public reuse libraries were in a situation similar to that of post-Soviet Poland. Confronted with a tiny market, they felt compelled to follow a strategy of becoming the dominant or even the sole provider. To achieve this goal, they attempted to provide all goods in a generalized manner to the largest possible fraction of the users. In short, they were all offering the same goods in nearly the same way.

The RIG's hope for the future is that the availability of the vast potential marketplace provided by the World Wide Web and standards for interoperability will motivate the major library operators and the independent entrepreneurs to continue their nascent efforts at differentiation and lead to a major industry in software reuse.

Further information about the RIG may be found at its Web site at `http://www.rig.org/`.

References

[1] *IEEE Standard for Information Technology - Software Reuse - Data Model for Reuse Library Interoperability: Basic Interoperability Data Model (BIDM).* IEEE Std 1420.1, 1995.

[2] R. F. Boisvert, S. E. Howe, and D. K. Kahaner. GAMS: A framework for the management of scientific software. *ACM Trans. Math. Softw.*, 11(4):313–355, Dec. 1985.

[3] R. F. Boisvert, J. L. Springmann, and M. L. Strawbridge. The GAMS virtual software repository. In *Proceedings of the Thirtieth Semi-Annual Meeting*, pages 68–72, Gaithersburg, MD, September 1992. Cray User Group, Five Point Editorial Services.

[4] S. Browne, J. Dongarra, S. Green, K. Moore, T. Pepin, T. Rowan, and R. Wade. Location-independent naming for virtual distributed software repositories. In *ACM SIGSOFT Symposium on Software Reusability*, Seattle, Washington, Apr. 1995.

[5] S. Browne, J. Dongarra, S. Green, K. Moore, T. Rowan, R. Wade, G. Fox, K. Hawick, K. Kennedy, J. Pool, R. Stevens, R. Olson, and T. Disz. The National HPCC Software Exchange. *IEEE Computational Science and Engineering*, 2(2):62–69, Summer 1995.

[6] D. Dikel. Reuse library interoperability group: Mission and purpose. In *Proceedings of the ACM Washington Ada Symposium*, July 1992.

[7] J. J. Dongarra and E. Grosse. Distribution of mathematical software via electronic mail. *Commun. ACM*, 30(5):403–407, May 1987.

[8] C. Lillie and J. Moore. Asset Source for Software Engineering Technology (ASSET) Software Reuse and Reengineering. In *NISQP Software Reuse and Re-engineering Conference*, Alexandria, Virginia, Apr. 1991.

[9] J. W. Moore. The impact of reuse library interoperability upon a software reuse industry. In *Sixth Workshop on Institutionalizing Software Reuse*, Owego, NY, Nov. 1993.

[10] J. W. Moore. The use of encryption to ensure the integrity of reusable software components. In *Third International Conference on Software Reusability*, Rio de Janeiro, Brazil, Nov. 1994.

[11] R. Rivest. The MD5 message-digest algorithm. *Internet Request for Comments*, 1321, Apr. 1992.

[12] J. Rumbaugh, M. Blaha, W. Premerlani, F. Eddy, and W. Lorensen. *Object-Oriented Modeling and Design.* Prentice-Hall, 1991.